From online texts to Landscape Character Assessment: Collecting and analysing first-person landscape perception computationally

Koblet, Olga ; Purves, Ross S

Abstract: Inspired by the narrative nature of Landscape Character Assessment (LCA), we present a complete workflow to (i) build a collection of almost 7000 online texts capturing first-person perception of the Lake District National Park in England, and (ii) analyse these for sight, sound and smell perception. We extract and classify more than 28,000 descriptions referring to sight, almost 1500 to sound and 78 to smell experiences using text analysis. The resulting descriptions can be explored for the whole Lake District revealing for example, how traffic noise intrudes on experiences in the mountains close to transportation axes. Linking descriptions to LCA areas allow us to compare properties of different regions in terms of scenicness or tranquillity at a macro-level by identifying, for example, LCA areas dominated by descriptions of tranquillity or anthropogenic sounds. At a micro-level, we can zoom in to individual descriptions and landscape elements to understand how particular places are experienced in context. Local experts gave positive feedback about the utility of such information as a monitoring tool complementary to existing approaches. Our method has potential for use both in allowing comparison over time and identifying emerging themes discussed in online texts. It provides a scalable way of collecting multiple perspectives from written text, however, more work is required to understand by whom, and why, these contributions are authored.

DOI: https://doi.org/10.1016/j.landurbplan.2020.103757

Originally published at:
Koblet, Olga; Purves, Ross S (2020). From online texts to Landscape Character Assessment: Collecting and analysing first-person landscape perception computationally. Landscape and Urban Planning, 197:103757.
DOI: https://doi.org/10.1016/j.landurbplan.2020.103757

Landscape and Urban Planning 197 (2020) 103757

Contents lists available at ScienceDirect

Landscape and Urban Planning

journal homepage: www.elsevier.com/locate/landurbplan

From online texts to Landscape Character Assessment: Collecting and analysing first-person landscape perception computationally

Olga Koblet*, Ross S. Purves

Department of Geography, University of Zurich, Switzerland

ARTICLE INFO

Keywords:
Landscape perception
LCA
Text analysis

ABSTRACT

Inspired by the narrative nature of Landscape Character Assessment (LCA), we present a complete workflow to (i) build a collection of almost 7000 online texts capturing first-person perception of the Lake District National Park in England, and (ii) analyse these for sight, sound and smell perception. We extract and classify more than 28,000 descriptions referring to sight, almost 1500 to sound and 78 to smell experiences using text analysis. The resulting descriptions can be explored for the whole Lake District revealing for example, how traffic noise intrudes on experiences in the mountains close to transportation axes. Linking descriptions to LCA areas allow us to compare properties of different regions in terms of scenicness or tranquillity at a macro-level by identifying, for example, LCA areas dominated by descriptions of tranquillity or anthropogenic sounds. At a micro-level, we can zoom in to individual descriptions and landscape elements to understand how particular places are experienced in context. Local experts gave positive feedback about the utility of such information as a monitoring tool complementary to existing approaches. Our method has potential for use both in allowing comparison over time and identifying emerging themes discussed in online texts. It provides a scalable way of collecting multiple perspectives from written text, however, more work is required to understand by whom, and why, these contributions are authored.

1. Introduction

According to the European Landscape Convention (Council of Europe, 2000) public perception should be taken into account in landscape assessment. However, in practice this is difficult (Jones and Stenseke, 2011). How do we collect the explicit opinions of people who have visited and experienced landscape? Methodologically, in-depth interviews and other qualitative approaches are one possibility, but they are typically applied only locally (Bieling, Plieninger, Pirker, & Vogl, 2014; Caspersen, 2009; Clemetsen, Krogh, & Thorén, 2011). Participatory GIS (PPGIS) and surveys are easier to use for larger areas, however, they often capture the views of local residents and exclude others interacting with landscapes (Brown & Reed, 2009; Bruns & Stemmer, 2018; Kienast, Frick, van Strien, & Hunziker, 2015). Therefore, a problem exists not only in sourcing public perception of landscapes, but also in collecting diverse voices (Butler, 2016). In this paper we combine the need to capture different groups and provide solutions suitable for large regions by collecting and computationally analysing texts describing a range of individual experiences in landscapes.

One pioneering framework in landscape assessment, initiated in the UK in the 1980s, and since adapted by many countries – Landscape Character Assessment (LCA) – aimed for a shift from describing iconic landscapes, to describing all landscapes without exception. An important goal was capturing properties distinguishing distinctive areas from their neighbours (Fairclough et al., 2018; Tudor, 2014). LCA's guidelines emphasise the importance of individual experiences in landscapes perceived through multiple senses "such as smell/scent, tranquillity, noise, and exposure to the elements (wind and rain for example)" (page 42, Tudor, 2014). The LCA process is divided into 4 steps: definition of purpose and scope, desk study, field study and final classification and description (Tudor, 2014). The desk phase collects information about physical properties of landscapes and delineates areas of distinctive character (LCA areas). Fieldwork is mostly concerned with *in situ* perception of people towards given landscapes. The results are then compiled into rich textual descriptions for each LCA area. Important challenges for LCA include integrating perspectives and perceptions from multiple people (and not only experts) and multiple senses (not overprivileging sight) (Swanwick & Fairclough, 2018). Furthermore, different groups of people value landscapes in different ways. For example, Butler (2016) adopted categories identified by Relph (1976) to a landscape context, and criticised the dominance of the 'objective outsider' in LCA. Considering other voices, and in

* Corresponding author.
E-mail addresses: olga.koblet@geo.uzh.ch (O. Koblet), ross.purves@geo.uzh.ch (R.S. Purves).

https://doi.org/10.1016/j.landurbplan.2020.103757
Received 10 July 2019; Received in revised form 7 January 2020; Accepted 19 January 2020

particular those of 'insiders' - should be part of the LCA process (Butler, 2016; Swanwick & Fairclough, 2018). These challenges are not unique to LCA and are relevant for all integrated approaches to landscape monitoring including experienced perception at some level: these may include what Kienast, Wartmann, and Hunziker (2019) term indicator-driven approaches and comprehensive narratives of landscapes.

One possible approach to addressing this gap is through the use of rich written sources, as has long been practised by environmental historians (Cronon, 1992). Galaz et al. (2010) and Daume, Albert, and von Gadow (2014) suggested using texts extracted from the internet in ecological monitoring and identifying unanticipated threats in forest monitoring respectively. Bieling (2014) demonstrated that short written stories can be used not only as a source to detect events and species, but also in the context of Cultural Ecosystem Services. The short texts contained information about spiritual and inspirational values of landscapes, concepts related to sense of place and identity, cultural heritage and aesthetics and have many parallels with sourcing perception in LCA. Wartmann, Acheson, and Purves (2018) compared 50 texts from online hiking blogs to free-listing interviews conducted *in situ* and tags submitted to the image hosting platform Flickr. The hiking blogs contained more information about sense of place, activities and perceptual landscape aspects than free-listing interviews and Flickr tags. If we can create a reproducible workflow, capable of collecting such short texts in an automatic way, we can potentially overcome the problem of time-intensive interviews for regions where landscape descriptions are available online. Furthermore, by using text, we can retain the advantages Bieling (2014) and Wartmann et al. (2018) identified in terms of rich narrative, but collect them for large areas. Finally, if the workflow is repeatable, we can also explore how landscape descriptions vary over time, a key task in monitoring.

In this work we explore sight, sound and smell perception as well as tranquillity. References to sight prevail in both oral and written accounts of English language (San Roque et al., 2015; Winter, Perlman, & Majid, 2018), and the importance of the ways sentiments towards landscapes are expressed through language has been debated since the Romantic era introduced notions such as 'sublime' and 'picturesque' landscapes (Donaldson, Gregory, & Taylor, 2017; Herlin, 2016).

Sounds present in landscapes are perceived selectively (Fisher, 1999) with 'no direct correlation between physical measurements of loudness and perceptions of noise' (page 641, Coates, 2005). Nonetheless, a taxonomy, developed in ecoacoustics is valuable since preferences for sounds vary according to perceived emitters. For example, though natural (e.g., waterfall) and anthropogenic (e.g., jet engine) sounds may have very similar signatures, preference is expressed as a function of the nature of perceived emitters (Fisher, 1999). Three classes of sound emitter are proposed: anthrophony (sounds produced by people), biophony (sounds of animals), and geophony (non-biological natural sounds) (Krause, 2008).

To these we add perceived tranquillity, which has been shown to be a combination of sight and sound (Carles, Barrio, & De Lucio, 1999; MacFarlane, Haggett, Fuller, Dunsford, & Carlisle, 2004; Pheasant, Horoshenkov, Watts, & Barrett, 2008). One common way to classify tranquillity uses a continuous scale from least to most tranquil landscapes (e.g., Hewlett, Harding, Munro, Terradillos, & Wilkinson, 2017). However, to capture ways tranquillity is written about, we developed a taxonomy (Chesnokova et al., 2019), with four classes: combination of sight and sound, contrasting sounds, no-movement and the class of silence and tranquil sounds. As for sounds, smells are often described through emitters, e.g., 'smell of birch' (Granö, 1997; Majid & Burenhult, 2014; Quercia & Schifanella, 2015) and can be similarly classified into anthropogenic sources, and those emitted by plants or animals.

Since explicitly collected short stories and short texts available online show high potential for eliciting public perception of landscapes, our aim is to demonstrate that large volumes of written texts, retrieved from the internet, are an effective source of information about public perception towards landscapes, specifically in the context of LCA. To approach this aim, we set out to investigate the following research questions:

RQ1: How can we build a spatial referenced corpus (collection of text documents) of first-person landscape perception?

RQ2: What sorts of perception do we find in our corpus, and from whom?

RQ3: How can these results be applied for LCA?

2. Methods

To illustrate our approach, we focus on a specific case study region, the English Lake District (2.1). Using this region as an example, we describe a general workflow to collect a corpus of documents from the web, containing first-person landscape perception in the Lake District (2.2). We then demonstrate how this corpus can be analysed, extracting and classifying descriptions of sights, sounds and smells experienced by the writers of this corpus (2.3). Finally, we associate descriptions with the region as a whole, LCA areas for the Lake District (Watkins, 2008) and points associated with individual landscape elements (2.4) (Fig. 1).

2.1. Case study region

To test the potential of written textual sources we selected an area of more than 2000 km^2 in the North-West of England – the Lake District National Park (Fig. 2) – established in 1951, which became a UNESCO World Heritage Site in 2017 (Nomination, 2017). This region is not only important because of its status as a National Park and World Heritage Site, but also because of its prominence in writing about landscape and nature in English. Multiple authors (e.g., Samuel Taylor Coleridge, Dorothy and William Wordsworth) celebrated the Lakes as a place of walking and nature appreciation in the Romantic Period at the start of the 19th century (Donaldson et al., 2017). This tradition of writing has continued to the current day and now also reflects the wide range of outdoor activities undertaken there. The area is characterised by rugged topography including England's highest mountain Scafell Pike (978 m) and its deepest and longest lakes (Wastwater (74 m) & Windermere (18 km)). In the 18th–19th century the Lake District became a centre of different types of industry, including quarrying of slate, limestone, and granite (Watkins, 2008).

2.2. Creating a corpus of first-person landscape perception in the Lake District

The internet as a whole was estimated at the time of writing of this paper to contain 5.86 billion documents (de Kunder, 2019; van den Bosch, Bogers, & de Kunder, 2016). This enormous volume of natural language clearly has great potential for analysis in multiple fields. However, before analysing landscape perception, we first need to identify thematically *and* spatially relevant texts: texts containing references to first-person landscape perception *in* the Lake District. Before building a corpus we identify three key requirements. The first of these is precision – the proportion of relevant descriptions should be as high as possible. The second is recall – as many relevant descriptions as possible should be returned. The third requirement, of particular importance if we are collecting individual experiences, is that we minimise the number of duplicate, or near duplicate documents. To maximise recall, we first used a set of search terms to programmatically retrieve candidate descriptions from search engine (2.2.1). We then increased precision on this initial document set by using machine learning to classify thematically relevant texts containing first-person landscape perception (2.2.2). Spatial precision was increased by a use of a spatial filter (2.2.3) before similar documents were removed (2.2.4).

2.2.1. Initial corpus

Our initial set of candidate documents was retrieved by a Python

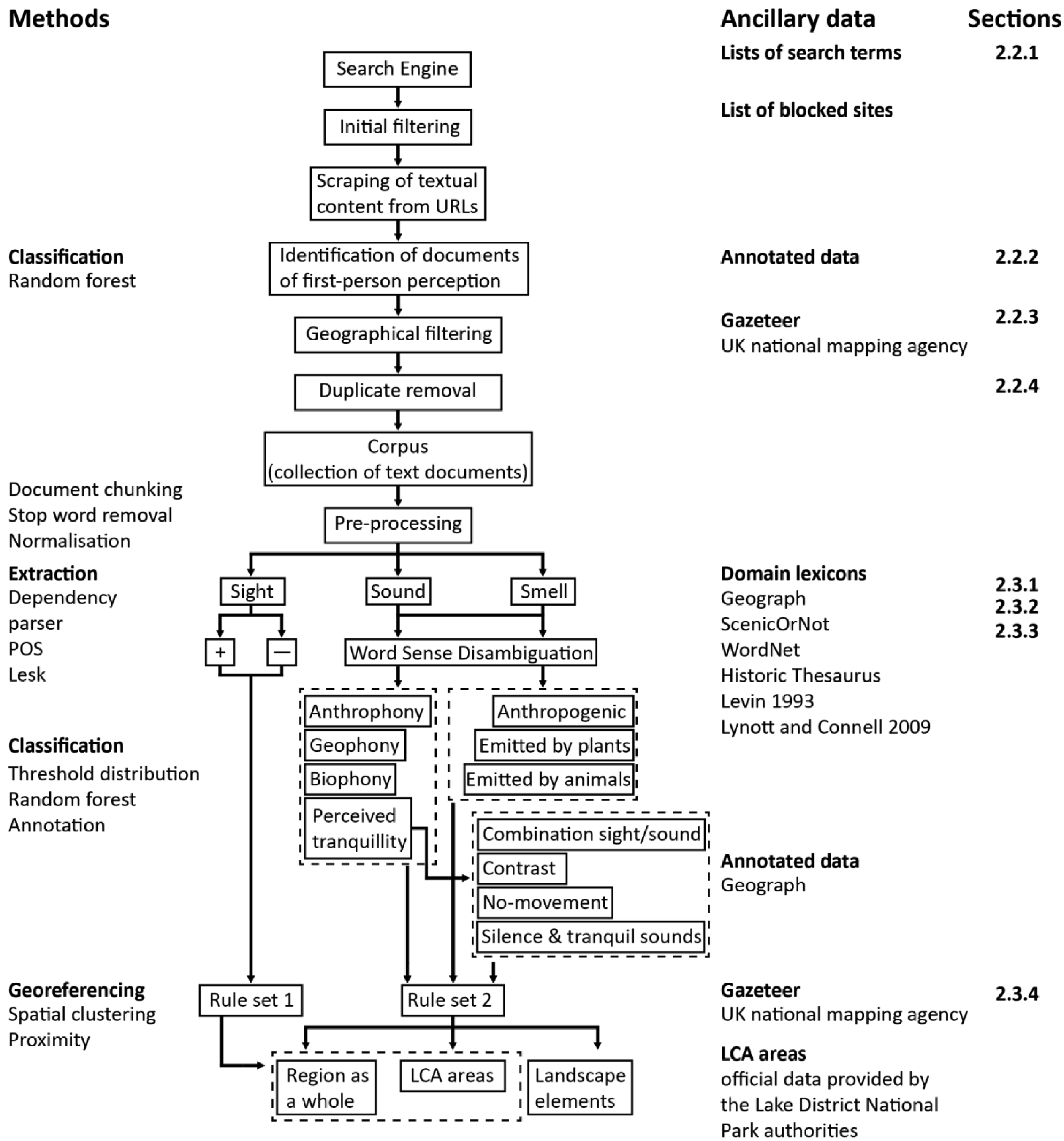


Fig. 1. Workflow including corpus creation and extraction, classification and georeferencing of first person descriptions of sights, sounds and smells.

program from the Bing search engine, using an application programming interface (API) to submit multiple queries. All queries were made with the market set to 'en-GB', specifying both preferred language and region of interest (Bing Web Search, 2019).

Each query consisted of a set of search terms likely to retrieve relevant documents (Joho & Sanderson, 2000). Initial experiments showed including "I" in the search terms increased the proportion of documents containing first-person perception. To retrieve documents relevant to the Lake District, and its landscape, we also used place names as search terms. The choice of names is central to the corpus which emerges (c.f. Davies, 2013; Wartmann et al., 2018), and we sought to address two of the categories suggested by Relph (1976) – 'behavioural insiders' and 'empathetic insiders'. 'Behavioural insiders' visit landscapes deliberately and visual patterns play a primary role. 'Empathetic insiders' do not just look at landscapes, but appreciate their identity through 'deliberate effort of perception' and understanding of 'place as rich in meaning' (page 54, Relph, 1976). To find descriptions written by 'behavioural insiders', we used a list of the names of 150 major outdoor attractions listed by TripAdvisor in the Lake District (c.f. Richards & Tunçer, 2018). These include architectural objects (e.g., castles and churches), historical landmarks, parks and gardens, viewpoints, waterfalls and houses of writers (see Appendix 1). To find descriptions of 'empathetic insiders' we used Wainwrights' list of Lake District summits, a particularly popular list for 'hill-bagging' (see Appendix 1). We assumed that those visiting such summits might more closely match the notion of 'empathetic insiders', since they may experience the landscape more intimately and more often, making many return trips to the region to collect all of the summits on the list. These lists of names can be substituted or expanded with other place names, depending on the nature of the case study region (e.g., using street

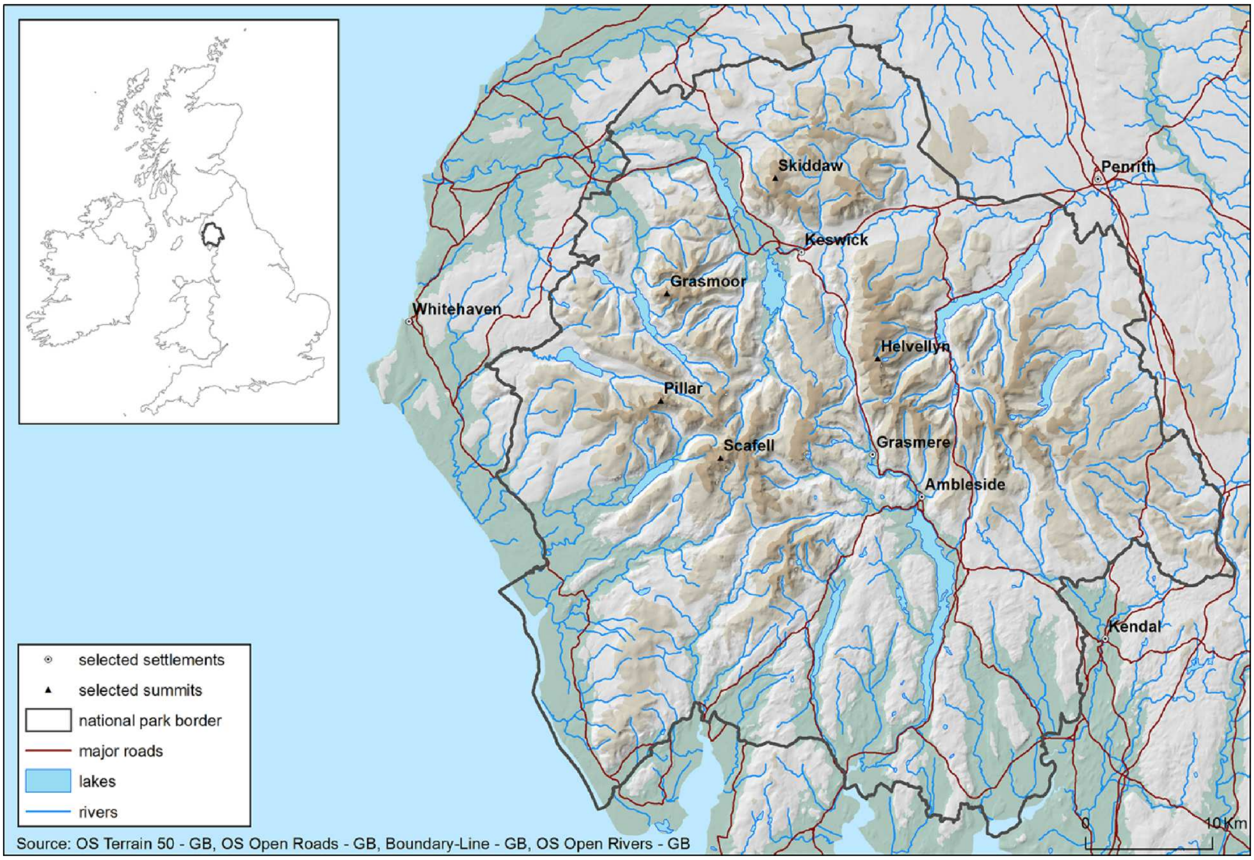


Fig. 2. The Lake District National Park and its topography.

names in an urban landscape). Place names are subject to both referent class ambiguity (e.g., Sail refers to a peak and is a common word in English) and reference ambiguity (e.g., Blencathra is also known as Saddleback) (Jones, Purves, Clough, & Joho, 2008). We dealt with referent class ambiguity by expanding our queries, adding 'Wainwright' to all searches for summits (c.f. Overell & Rüger, 2008). Reference ambiguity was dealt with by adding all known names for a given location to the search terms.

Each query returned a list of web addresses (known as URLs). Before analysing the content associated with URLs, we used a manually compiled list to programmatically remove those unlikely to contain first-person descriptions of landscape. These included words related to local government, accommodation, and Wikipedia pages, since these often contain local place names but not descriptions of individual experiences (e.g., 'gov.uk' as in https://www.lakedistrict.gov.uk/) (see Appendix). We also removed all duplicate URLs. From the remaining URLs we scraped visible textual content of all webpage elements using the Python library 'scrapy', excluding headers, footers, sidebars, and comments (Fig. 3) in accordance with the web-scraping policies of individual web sites (Greenaway, 2017; Lawson, 2015).

2.2.2. Classifying thematically relevant documents

Our search terms were designed to return documents likely to include first-person landscape perception, but a second classification step was necessary to remove false positives and increase precision. To do so, we applied a random forest, a supervised machine learning classifier well suited to textual features, using Python library 'scikit-learn' (Criminisi, Shotton, & Konukoglu, 2011; Pedregosa et al., 2011). We manually annotated training data in a preliminary study (Chesnokova and Purves, 2018a) and trained the classifier using three groups of features: the 250 most frequent words, presence of selected personal pronouns and the 50 adjectives and nouns most frequent in relevant/not relevant descriptions respectively. We split 641 annotated texts (see annotation rules in Appendix Table 1) into a 50% training and 50% test sets, and achieved precision of 0.84.

2.2.3. Filtering spatially irrelevant texts

Although we retrieved URLs using place names, these were not necessarily found in the scraped text we extracted for analysis (c.f. Fig. 3 where Rome occurs in the sidebar but not in the scraped text). We therefore performed a simple toponym recognition step using the complete list of place names used as search terms (c.f. 2.2.1) and a place name gazetteer from the UK national mapping agency for the Lake District. To account for small differences in spelling we used Levenshtein distance as implemented in Python library 'Fuzzy String Matching' (Arias, 2019), a string metric which measures how many characters need to be inserted, deleted or substituted to move from one string to another (e.g., the Levenshtein distance between cat and cars is 2 since we substitute r for t and insert s). We also used simple heuristics to match potentially compoundable nouns (e.g., Derwentwater/Derwent Water/Derwent water).

2.2.4. Eliminate similar documents

Finding duplicate descriptions contributed via different URLs is an important step, as we do not want to emphasise landscape characteristics found in multiple texts with the same source. We filtered out all descriptions with an overall string similarity of more than 80% (Python library 'Fuzzy String Matching') to create our final corpus ((Zachara & Palka, 2016; Arias, 2019; Gonzalez and Rodrigues, 2017)).

2.3. Extracting and classifying descriptions of sights, sounds and smells from our corpus

Having created a corpus of first-person landscape descriptions, we

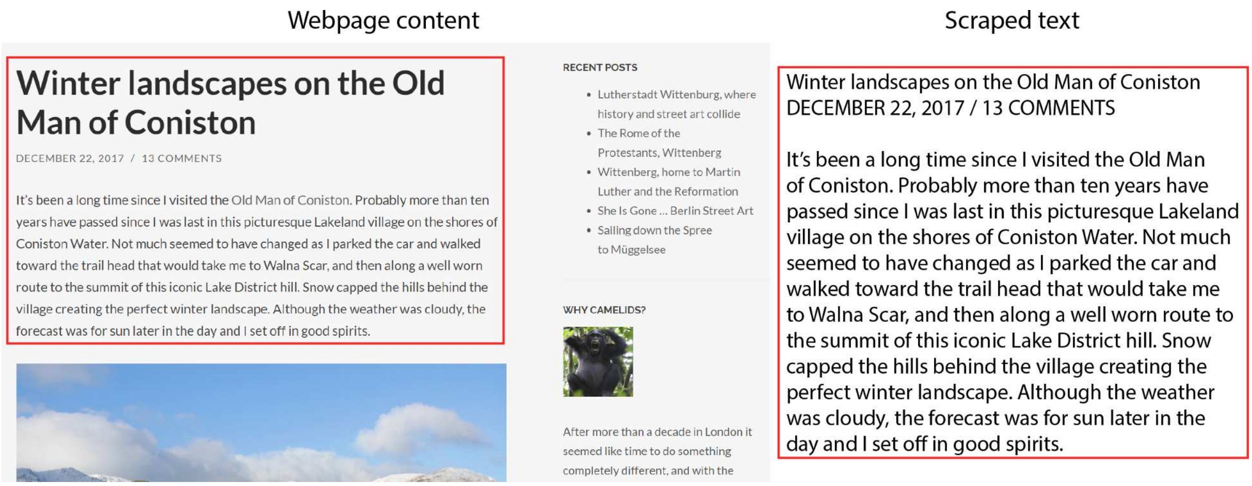


Fig. 3. Original webpage content on the left and the scraped textual content without sidebars on the right (https://notesfromcamelidcountry.net/category/coniston/, Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported Licence).

analysed experiences of sights, sounds and smells in the Lake District.

For all senses, we first performed a range of natural language preprocessing steps using the Python library 'spacy' (Srinivasa-Desikan, 2018). These included detecting paragraphs, sentences and words in documents, part of speech tagging (e.g., identifying adjectives, verbs and nouns), removal of stop words (e.g., 'a', 'the'), normalisation of words to lower case and extraction of lemmas (dictionary roots of words with the same semantic meaning) (Manning & Schutze, 1999).

After preprocessing, we automatically extracted sub-corpora containing references to sights, sounds and smells respectively. For extraction we used lexicons (lists of domain relevant terms) and pattern matching combining for example lemmas and parts of speech. For smells and sounds, we also performed word sense disambiguation, removing irrelevant uses of words (e.g., sound can refer to a body of open water).

We then further classified the extracted descriptions, in terms of scenic or unattractive elements of the visual landscape, classes of tranquillity as identified through references to sounds, and emitters for sounds and smells using a combination of machine learning and manual annotation.

2.3.1. *Extracting and classifying sights in the landscape*

References to visual perception are common in language and use a wider range of words in English than other senses (San Roque et al., 2015; Winter et al., 2018). Sentiment is often conveyed through phrases combining adjectives with nouns (e.g., compare *overcrowded summit* with *beautiful lake*), and we used this observation to guide our analysis (Liu, 2012). We were interested in collecting particularly negatively or positively connoted descriptions associated with visual perception from our corpus. This task is associated with sentiment analysis in natural language processing where lexicons are used to identify words or phrases found in, for example, positive or negative reviews (Kaji & Kitsuregawa, 2007; Lu, Castellanos, Dayal, & Zhai, 2011). Such a lexicon does not exist for landscape. To create one, we needed a collection of ratings related to landscape and texts associated with those ratings. The ScenicOrNot project (http://scenicornot.datasciencelab.co.uk/) collected more than 220,000 ratings of "scenicness" (with values between 1 and 10) for images from the Geograph project (http://www.geograph.org.uk/), a collection of representative pictures and descriptions for the whole of the UK. ScenicOrNot ratings are available under an Open Database Licence and the Geograph dataset under a Creative Commons Licence.

To build our lexicon we relied on three observations. First, since we have ratings for individual pictures and their descriptions, we can associate phrases with scenic or unattractive landscapes. Second, the Lake District is valued for its scenicness, and thus we expect unattractive descriptions to be rarer than in the UK as whole. Third, since we also know the overall distribution of scenicness ratings, we can identify phrases which are used particularly often to refer to unattractive or scenic landscapes.

Based on these observations we collected descriptions associated with scenic (mean scenicness + 2 standard deviations) and unattractive descriptions (mean scenicness − 1 standard deviation) (Fig. 4). Doing so resulted in 4847 scenic and 26,029 unattractive descriptions for the UK as a whole.

To create our lexicon of phrases associated with unattractive and scenic landscapes, we then extracted adjectival modifiers using a dependency parser (e.g., from the phrase, 'stunning panoramic views', we extracted two pairs: 'stunning views' and 'panoramic views') (Honnibal & Johnson, 2015), and tested these for significance compared to all descriptions (Chi-square test, df $= 1$, $p < 0.005$). We only retained phrases which were associated with particularly high or low ratings of scenicness and not simply common overall. The resulting lexicon contained 184 scenic phrases and 214 associated with unattractive descriptions (see Appendix).

2.3.2. *Extracting and classifying sounds in the landscape*

To extract descriptions related to sounds in the landscape, we also used a lexicon. We took a top-down, knowledge-based approach, and

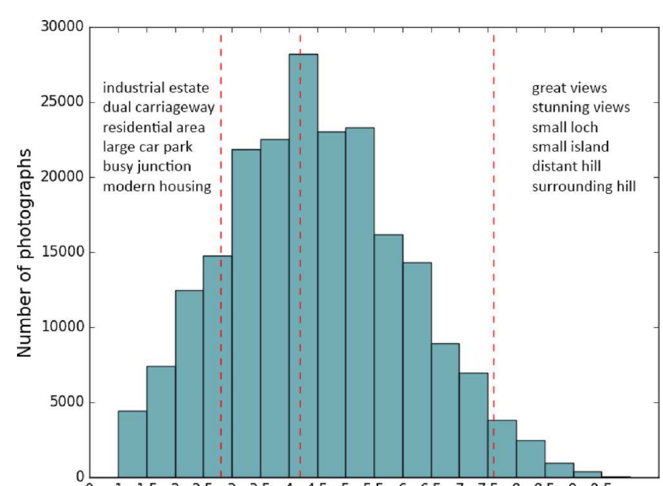


Fig. 4. Distribution of scenicness values for all pictures and descriptions, thresholds for scenic and unattractive descriptions and examples of adjectival modifier pairs extracted.

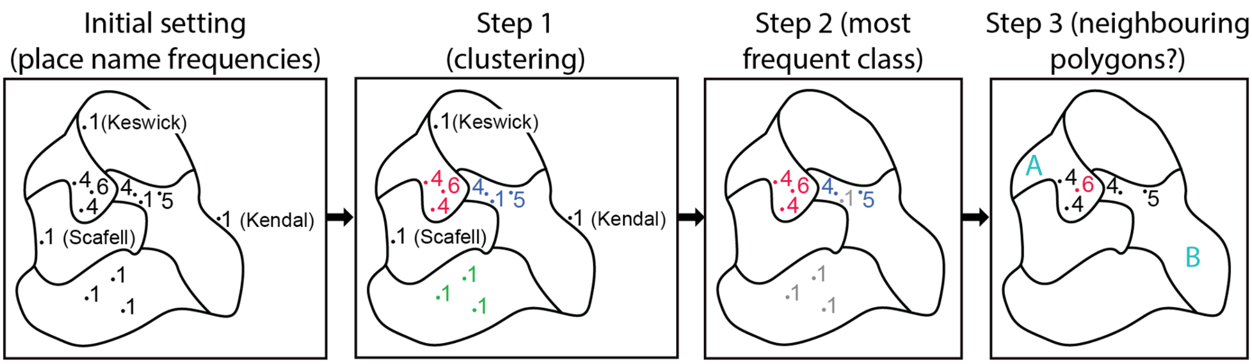


Fig. 5. The workflow to assign documents to LCA areas. Different colours in Step 1 correspond to different clusters, blue and red locations in Step 2 belong to the most frequent class, and red location with frequency 6 in Step 3 is the most frequent place name. Areas A and B are LCA areas associated with the description. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

built a lexicon consisting of English verbs of sound emission, sound existence and sounds made by animals (Levin, 1993), and synonyms for all these verbs according to WordNet (Fellbaum, 1998). We added a list of adjectives related to sounds. Since our previous work had shown that silence is an important sound related quality in the landscape (Chesnokova et al., 2019), we added a list of terms referring not only to perceived absence of sound, but to tranquillity as a holistic combination of sight and sound experiences. We took terms from the Historical Thesaurus of English (https://ht.ac.uk/) in the categories "inaudibility," "faintness/weakness" and "quietness/ tranquillity" to create the final lexicon consisting of 262 sound and tranquillity related words (see Appendix).

We then extracted candidate descriptions containing sound-related first-person perception by matching lemmas of lexicon terms to our corpus. This generated many false positives, since words describing sound experiences are highly polysemous (e.g., echo can be used literally with respect to sound or figuratively to describe styles). We disambiguated such cases by, first, controlling for the correct sense of verbs and nouns using WordNet categories as implemented in the Lesk algorithm (Manning & Schutze, 1999). Second, we developed rules using part of speech tagging (e.g., taking into account 'still' only if it is labelled as adjective as in 'still waters' and not as adverb as in 'still hungry'). Finally, based on examination of false positives, we added additional rules to filter common ambiguities (e.g., removing 'screaming calves', usually referring to muscular pain).

We used an existing taxonomy of sound emitters, classifying sounds as biophony, anthrophony and geophony (Chesnokova and Purves, 2018b), adding an additional class often found in first-person landscape descriptions, absence of sound, to account for descriptions conveying a general sense of peace in terms of sounds and sights (c.f. Pheasant & Watts, 2015). We manually annotated 8784 descriptions in the Geograph collection according to this taxonomy with a Cohen's Kappa inter-annotator agreement of 0.88 (Landis & Koch, 1977). After exploring these texts in detail, we redefined absence of sound to include perceived tranquillity (see Appendix, Chesnokova and Purves, 2018b).

Using these annotated data, we implemented a random forest classifier using Python library 'scikit-learn' (Criminisi et al., 2011; Pedregosa et al., 2011), using as features the 500 most common words, a list of British birds and mammals and a list of natural elements and related qualities to classify extracted and disambiguated sound descriptions as either biophony, anthrophony, geophony or tranquillity (see Appendix), and achieved precision of 0.81.

Since tranquillity can be usefully and reliably classified by human annotators, we sub-classified each description referring to tranquillity to one of the following 4 categories: contrasting sounds; combination of sight and sound perception; no-movement and total silence and tranquil sounds (Chesnokova et al., 2019) (see annotation rules in Appendix Table 2).

2.3.3. Extracting and classifying smells in the landscape

The final sense extracted from our corpus was smells. As for sound, we created a lexicon based on verbs of smell emission (Levin, 1993) extended by WordNet lists of olfactory categories and adjectives with dominant modality "olfactory" (Lynott & Connell, 2009). This lexicon contained 29 words (see Appendix). We disambiguated candidate descriptions using an analogous process to that performed for sound. Finally, since references to smells were relatively rare, we classified these manually into those emitted by plants, animals and anthropogenically.

2.4. Associating classified descriptions with space

Having extracted and classified descriptions associated with sights, sounds and smells, we were left with a subset of documents containing relevant descriptions of individual senses. For each of these descriptions, we could identify the sentence related to landscape perception and an attribute indicating associations with senses and the resulting classification. An individual description could be associated with one or more senses.

These descriptions could be analysed without any further processing, as characterising the Lake District. However, our motivation was to provide descriptions relevant to LCA, and to do so we had to explicitly link text to space. We used place names found in the texts to link sight related descriptions to LCA areas (Watkins, 2008) and sounds and smells to the point locations of place names (e.g., summits or settlements) found in their descriptions.

To assign sight-related documents to LCA areas we performed three steps having initially calculated place name frequency in each text (Fig. 5). First, we applied density-based clustering as implemented in PostGIS (Moncla, Gaio, & Mustière, 2014; PostGIS, 2019) to disambiguate and detect outliers of seen but not visited locations. Second, we created three classes of place name frequency based on Jenks natural breaks data clustering (Dara-Abrams, 2011) and retained only the most frequent class (which we assume to be more likely to be visited and thus experienced). Finally, we took the most frequent place name and performed a region-based disambiguation on the other place names found in the most frequent class. All steps related to spatial analysis were performed using Python library 'arcpy' (Esri, 2019).

For sounds and smells, we first looked for a place name in the relevant sentence, checking for referent ambiguity (does this place name occur more than once in the Lake District?). If not, then we assigned the coordinates found in a gazetteer. In cases of referent ambiguity, we disambiguated using other place names found nearby in the text using a distance-based measure (Leidner, Sinclair, & Webber, 2003).

Finally, to reduce the effects of bias induced by participation inequality (Nielsen, 2006), we retained only one description if several had the same class and location and were generated by the same user.

Having performed these steps, we have a list of perceived landscape

Table 1
Corpora statistics.

	TripAdvisor	Wainwright
Number of search terms (including different spellings)	92	214 (233)
Initial number of extracted texts	13,110	34,150
Number of relevant texts in the Lake District	961	5909
Average paragraphs per text	49	79
Average sentences per text	81	104
Average words per text	1277	1120

properties associated with LCA areas (sights) and individual landscape features (sounds and smells).

3. Results and interpretation

3.1. Thematic corpus of the Lake District

Using our customised lists, we initially retrieved 13,110 and 34,150 texts, for search terms derived from TripAdvisor locations and Wainwrights' list respectively. After the filtering stages described in Fig. 1 we were left with a total of 6870 relevant texts and a corpus consisting of almost 8 million words (Table 1). Documents varied in the nature of the information they contained ranging from descriptions of participation in a single event (e.g., a hike to a summit) through multiple descriptions of different locations by the same individual or collections of descriptions of a single location from multiple users. Thematically, contrary to our expectations, both sets of texts contained a broad mix of activities and narrative types without a clear distinction between 'behavioural insiders' and 'empathetic insiders' (Relph, 1976) and we treated texts thereafter as a single corpus of first-person perception (Table 1).

To investigate the efficacy of our lists in retrieving spatially relevant texts, we report here on our ability to link documents to LCA areas. For the 71 LCA areas in the Lake District, we could collect more than 10 texts for 54. Only a single area – Lyth Valley – has no texts. Peripheral areas, and in particular the southern part of the Lake District, not described in Wainwright's list, have fewer texts. Unsurprisingly, more texts are found for famous parts of the region, containing both the high mountains of Scafell and the popular valley landscapes around Grasmere (Fig. 6).

3.2. Characterisation of perceived landscape properties

An important challenge in the development of new methods to extract information is their potential utility for practical applications. The challenge of assessing whether or not we extract useful information is well-known, and we are interested in identifying *interesting patterns*. These are characterised in the Knowledge Discovery domain by unexpectedness (a user learns something new) and actionability (a user learns something upon they might act) (Silberschatz & Tuzhilin, 1996). To evaluate these, and other properties of our results, we created a set of web maps which we used in subsequent interviews with experts in the Lake District (3.3). For sight, we generated word clouds for the 34 LCA areas with two or more texts per km^2 (c.f. Fig. 6) and scaled the top 50 scenic and unattractive adjective modifier pairs using spatial term frequency/inverse document frequency (Rattenbury & Naaman, 2009) to identify locally distinctive pairs (Fig. 7). For sounds and smells we visualised individual descriptions as points with extracted sentences, paragraphs and the original URL available (Fig. 7).

In the following, we give selected examples of how these maps can be used to characterise our region of interest – the Lake District – as a whole and at the spatial scale of the LCA areas. In doing so we move from macro- to micro-analysis by looking at emerging patterns of automatic analysis in the former and zooming in to interpret individual

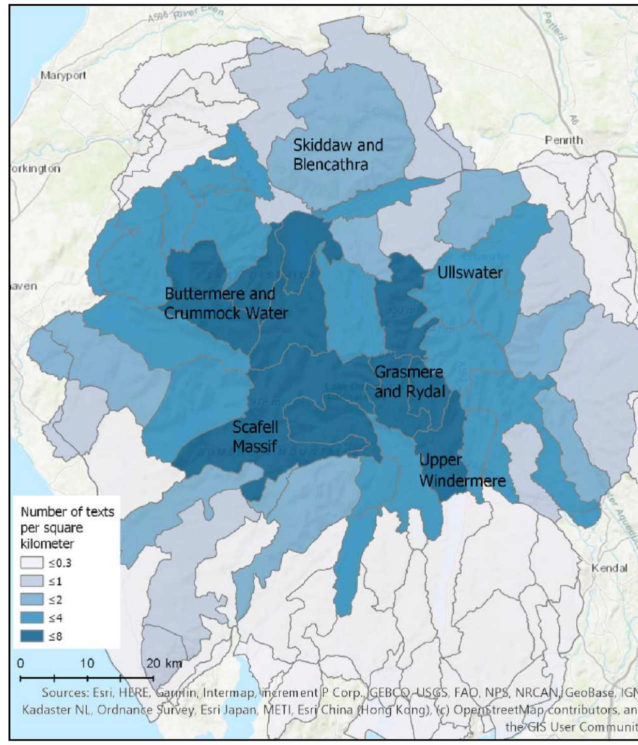


Fig. 6. Document density for LCA areas.

descriptions to better understand these patterns in the latter (Jockers, 2013).

3.2.1. Characterisation of perceived landscape properties on the scale of the Lake District

Using our domain specific scenicness lexicon, we extracted a total of 28,179 descriptions referring to scenic landscapes and only 266 descriptions referring to unattractive locations. These values are in themselves not surprising, since, first, our lexicon is based on the whole of the UK, and, second, the Lake District is characterised by its outstanding scenic qualities. Table 2 lists the ten most commonly occurring pairs found.

Many of the terms associated with scenic locations relate to generic visual properties such as great, good and stunning view(s). More experiential perception related to locomotion in the landscape is also common, as for example, steep ascent, steep descent and good path. These can be related to Wainwright, who considered 'bodily experience' an important component of landscape perception (Palmer & Brady, 2007). Zooming in to individual descriptions demonstrates the importance of this dimension for the writers in our corpus: "When we finally met up with the Stake Pass and could head down hill on a *good* solid and visible *path*, it was truly time to celebrate." (https://ramblingman.org.uk/walks/wainwrights/southernfells/esk_pike). Although unattractive elements are uncommon, they are also revealing, relating to negative sentiments towards transport (e.g., large (car) park, parked cars and dual carriageway): "Tarn Howes is a special place, albeit too close to a *large* car *park* and not seen at its best because of the poor light and lingering mist." (https://lonewalker.net/walkinfo.php?walk=412). Other references to unattractive elements refer to previous industry (e.g., old works, old machinery): "Dotted all around are spoil heaps, rusting iron cables lie along the path, bits of *old machinery* lay abandoned on the mountainside, and a metal tower from an aerial tramway lays toppled on its side." (https://notesfromcamelidcountry.net/category/coniston/). Such abandoned mines and quarries are common in the Lake District, however, they are often ignored by the writers of our corpora, following the tradition of William Wordsworth,

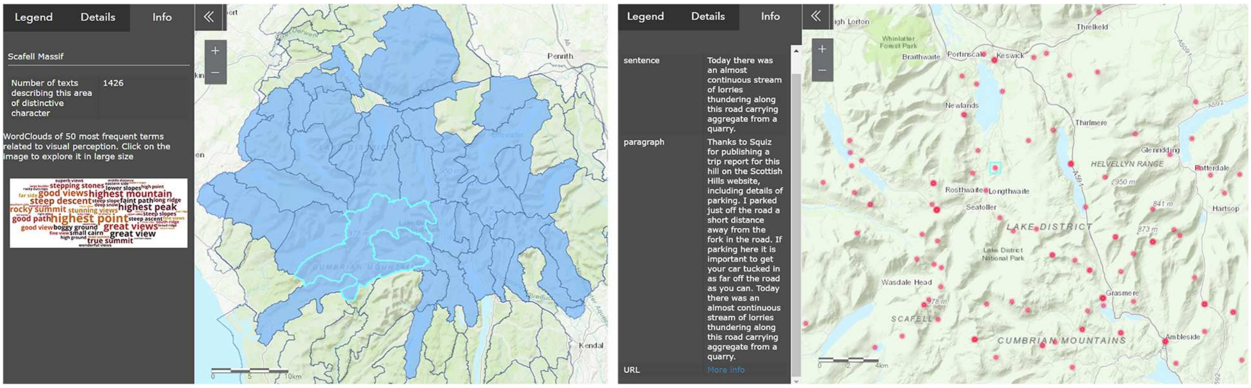


Fig. 7. Interface of the web maps demonstrating the results of our approach (tinyurl.com/LakeDistrictPerception). Left: Word cloud displayed for the selected LCA area "Scafell Massif". Right: Map of sound descriptions classified as anthrophony and an example description.

Table 2
Ten most frequent combinations from scenic and unattractive lexicons found in our corpus.

	Scenic pairs	Count	Unattractive pairs	Count
1	great views	1012	large (car) park	39
2	highest point	881	parked cars	26
3	good views	707	dual carriageway	18
4	steep descent	528	old works	12
5	steep ascent	370	old machinery	10
6	good path	353	adjacent park	6
7	good view	315	local shops	6
8	stunning views	314	static caravans	6
9	great view	306	much traffic	6
10	lower slopes	296	second bridge	5

Table 3
Summary of extracted descriptions of sound experiences per class.

Type of sound experience	Count
Combination of sight and sound perception	485
Contrasting sounds	275
No-movement	66
Tranquil sounds and total silence	60
Total perceived tranquillity	**886**
Anthrophony	174
Biophony	142
Geophony	278
Total assigned to emitter	**594**

who deliberately overlooked not only the appearance of – at this time still functioning – mines, but also the sound of the excavations (Taylor, 2018). Indeed, the majority of the sound related descriptions (886 out of total 1480 descriptions) refer to perceived tranquillity and the general absence of sound (Table 3).

Most common amongst such descriptions are those referring to a combination of sight and sound perception as in "The walk back to Wasdale Head is largely along the road but it is fairly quiet and offers ample opportunities to leave it to admire the views." (http://allthegearbutnoidea.blogspot.com/2018/06/a-circuit-of-wast-water.html). Contrasting sounds, which typically characterise a location favourably by comparison to, or despite, a nearby place's properties are also prominent as in this description of Broadcrag Tarn: "When the summit of Scafell Pike is crowded with excited and chattering groups of walkers, it is a place to visit for here very few people tread and apart from the occasional curious Herdwick sheep you are unlikely to meet a soul." (https://herdwickcountry.wordpress.com/2014/07/08/broadcrag-tarn/). These results reveal an important property of our work: having classified descriptions we can also perform micro-analysis to start to understand how the Lake District is experienced. In the class no-movement, the influence of the Romantic poets continues to persist, with its emphasis on the mirror-like reflection of lakes and tarns as in "We made a slight detour which brought us closer to Hayeswater, which was now free from mist and mirror like in its stillness." (https://these8boots.wordpress.com/category/patterdale/). These descriptions can be seen as not properties of individual locations, but rather an important part of the eponymous identity of the Lake District related to the water of its lakes and tarns. The influence of Wainwright's writing is also reflected, his favourite places Haystacks and Innominate Tarn appear in several descriptions of tranquillity, e.g., "In absolute silence we were treated to the reflection of the sky and the distant heights of Pillar in the totally flat waters of the tarn" (http://adventuresforthecommonman.blogspot.com/2018/04/sometimes-its-about-journey.html). As well as analysing tranquillity, we also produced maps of specific sound emitters: for anthrophony, rumbling traffic/traffic noise and for biophony, roaring/bellowing stags (Fig. 8). These demonstrate the relationship between different experiences and the Lake District as a whole, as well as specific locations. Thus, traffic noise is experienced and written about not only in the valley, but also on summits close to the major transportation axis of the A591, as for example on Helm Crag, "As you pass around to the other side of the fell the road noise from the A591 is more audible." (http://mydadsboots.blogspot.com/2015/10/the-lake-district-helm-crag-and-gibson.html). A distinct soundscape emerges from our texts as a cluster of locations south of Ullswater characterised by roaring stags: "This is the oldest red stag area in England and the "Rut" had began as roaring stags could be heard all around." (http://frasermackay.blogspot.com/).

Though smell perception is rare in English in comparison with sight and sound perception (Majid & Burenhult, 2014), smellscapes are still important in defining the unique character of places (Dann & Jacobsen, 2003). From our corpus, we identified 78 smell descriptions, 48 of which describe smells emitted by plants, 21 by anthropogenic sources (e.g., food, smoke) and 7 by animals. Anthropogenic descriptions capture popular tourist locations such as Grasmere's gingerbread shop, but also the common smell of burning brakes on the steep Hardknott pass: "Cars on 3 wheels coming around the steep hairpins, stinking of burning clutches and brake discs." (https://babtestingground.wordpress.com/2012/10/). We find dotted around landscape references to the scent of blossoming heather, sweet gorse and the pungent smell of dead sheep, often near steep Lakeland cliffs. As for sound perception, smell experiences reflect both temporally dynamic (e.g., blossoming flowers) and spatially variable (e.g., the final resting places of the unfortunate dead sheep) processes, but also encode information about the affordances of particular landscape elements (such as the heather covered slopes of Blencathra).

3.2.2. Relating perceived landscape properties to LCA

LCA has, at its core, the production of narrative descriptions related

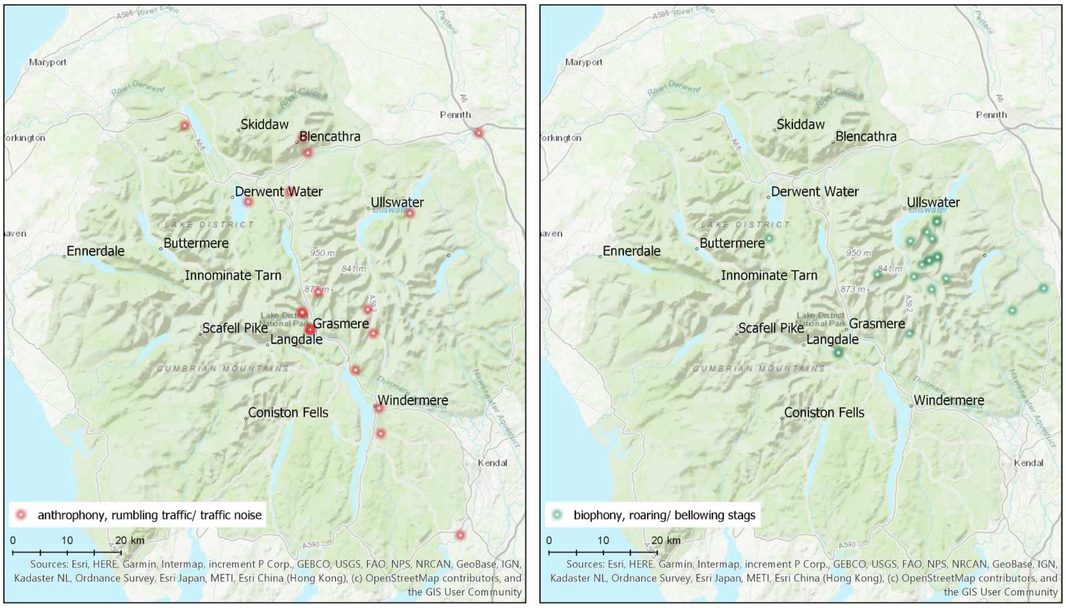


Fig. 8. Selected examples of sound experiences spatial distribution. Left: anthrophony, rumbling traffic/traffic noise, 19 descriptions. Right: biophony, roaring/bellowing stags, 20 descriptions.

to defined areas. Our extracted texts provide insights into perceived elements of sight, sounds and smells, making it possible to characterise landscapes at the level of the LCA areas, identify areas having similar characteristics in the region as a whole and compare them to their neighbours.

Our first macro-analysis exploration reveals that only 13 out of 34 LCA areas contain pairs from the unattractive lexicon. Two neighbouring areas: 'Helvellyn Range' and 'Brother's Water and Hartsop' are characterised by 'large (car) park', however, here the sentiments are not negative as in the example of Tarn Hows (c.f. 3.2.1), but reveal a more complex interplay between unattractive sight perception and overall positive sentiment: "Seventeenth century Hartsop village […] a lovely little place playing host to a rather *large* car *park*, an ideal starting point for the ascent of High Street." (http://www.one-foot.com/Over%20High%20Street%20return%20through%20pasture%20Beck%20Bottom%202012.html) and "The Helvellyn range is well served by a *large* car *park* in Glenridding" (http://allthegearbutnoidea.blogspot.com/2013/07/helvellyn-via-striding-edge.html).

For sound perception we calculated the proportion of texts describing sound experiences (in general and per class) to the total number of texts describing the corresponding area (Table 4). Looking at both sight and sound demonstrates that, for example, 'Upper Windermere' area has not only unattractive pairs, capturing its urban characteristics such as busy carriageway, parked cars and modern estate, but it also has the highest proportion of anthrophony related to traffic and noise referring to Royal Air Force training activities: "RAF trainer buzzing Kirkstone Pass" (http://www.loweswatercam.co.uk/130219_To_Sweden_with_Two_Pikes.htm). 'Claife Heights and Latterbarrow' area also contains several unattractive pairs, but sound is not characterised by anthrophony, but rather by its absence through contrasting sounds (13 of 23 tranquillity descriptions): "I have been through Kirkstone Quarries before and it is usually a hive of activity, but today it is uncharacteristically quiet." (http://www.flamingonion.co.uk/langdale_walk/). This text was written in 2012. In 2014 another author writes that the quarries have closed and "Now the slate cutting rooms and showrooms stand quiet and empty." (http://tandjinlakes2014.blogspot.com/), showing the potential of our approach to document change.

Two neighbouring areas 'Skiddaw and Blencathra' and 'Keswick and Derwent Water' are dominated by tranquillity related sound experiences (Table 4). However, in the 'Keswick and Derwent Water' area total silence is a scarce resource (only 1 mention) in comparison to 12 reports for 'Skiddaw and Blencathra'. 'Keswick and Derwent Water' is additionally characterised by high proportions of all other types of sounds: anthrophony, geophony and biophony. Zooming in we see that in contrast to 'Upper Windermere', with its traffic noise, anthrophony for 'Keswick and Derwent Water' is characterised by chugging boats and noises of other visitors. Two important tourist locations Lodore Falls and Ashness Bridge are within the borders of this area giving "sweet sound" (https://insearchofbritain.wordpress.com/tag/robert-southey/) and "wonderful sound" (https://upnoutside.wordpress.com/tag/ashness-bridge/), respectively. We also find references to pleasant sounds associated with wildlife (biophony), above all in the form of birds.

Despite the relative rarity of descriptions of smell perception, these can suggest important properties of the area as a whole, as for 'Skiddaw and Blencathra' where 4 of 9 descriptions mention the scent of "heather in full bloom" (https://www.wainwrightwalking.co.uk/ullock-pike-to-dodd/). Contrasting properties of the LCA areas to their neighbours confirms again the anthropogenic nature of 'Upper Windermere' with a single description referring to the anthropogenic smell of fish and chips, while the neighbouring area of 'Grasmere and Derwent Water' has seven descriptions, with four relating to a diverse range of plants including juniper, magnolia and hyacinths.

3.3. Expert group discussion

To evaluate the potential of our methods and its results, we visited the Lake District National Park authority for an expert discussion with, on the one hand, the authority itself, and on the other an important local lobby group (the Friends of the Lake District). We prepared a short presentation to introduce our approach, the web maps described above (tinyurl.com/LakeDistrictPerception) and a structured set of questions to discuss the utility of our approach based around a SWOT analysis (Strengths, Weaknesses, Opportunities and Threats). This study was not designed to provide a comprehensive evaluation, but rather feedback as to potential in a practical setting. Three participants took part in the expert discussion, and we summarise their feedback in Table 5.

A number of important points emerged from these discussions. The expert group explicitly saw the potential utility of such narrative descriptions for LCA, and also noted the value of a repeatable method for

Table 4
Five highest numbers of descriptions per type of sound emitter normalised by the number of texts describing corresponding area (in the brackets the absolute frequency is given).

	All sound experiences	Tranquillity and absence of sound	anthrophony	geophony	biophony
1	Skiddaw and Blencathra (117)	Skiddaw and Blencathra (82)	Upper Windermere (8)	Ennerdale (17)	Shap and Birkbeck fells (6)
2	Keswick and Derwent Water (74)	Claife Heights and Latterbarrow (23)	Keswick and Derwent Water (8)	Coniston Fells (13)	Martindale (8)
3	Coniston Fells (53)	Coniston Fells (33)	Borrowdale (11)	Keswick and Derwent Water (16)	Brother's Water and Hartsop (12)
4	Claife Heights and Latterbarrow (34)	Keswick and Derwent Water (42)	Thirlmere (6)	Borrowdale (22)	Keswick and Derwent Water (8)
5	Ennerdale (50)	Buttermere and Crummock Water (54)	Ullswater (10)	Grasmere and Rydal (18)	Kentmere Fells (9)

gathering such data. The importance of including detailed temporal information for monitoring was a key weakness, and a challenge that we discuss more below. In general, when working with subjective information, the participants pointed out the importance of context with respect to who had provided information, something also suggested as a weakness of current (expert-dominated) approaches to LCA (Butler, 2016). However, our experts also identified actionable patterns which we had both expected (relating landscape qualities more directly to perception) and which surprised us (identifying important indicator species not perceived by visitors). The potential utility of our tools in management, reflects broader initiatives in better understanding visitor behaviour in protected areas through such novel sources (Toivonen et al., 2019).

4. Discussion

In the introduction we set out aims, which can be summarised as methodological (how can we build a spatial referenced corpus of first-person landscape perception?), thematic (what sorts of perception do we find in our corpus, and from whom?) and potential (how can these results be applied for LCA?). We now discuss each of these questions in turn, pointing out not only strengths and weaknesses of our approach, but also more general implications for studies of landscape.

Edwards (2018) suggested that creative writing 'on, and better still *in* a landscape' (page 666) makes people focus on the senses and feelings these landscapes evoke, and that by sharing these personal stories people demonstrate their care for a particular landscape. She argues for inclusion of creative writing practices in the process of LCA. In this work we join her call, developing a customisable and repeatable workflow, to collect descriptions of first-person landscape perception, which we see as creative writing contributions published online. We used these texts to look at the ways LCA can capture multiple voices and become less expert-dominated.

We aimed for high precision (i.e., the descriptions we identify are likely to truly contain first-person perception) at the cost of missing other, potentially relevant descriptions. By using existing search engine APIs and lists of search terms we can rapidly build, filter and extract descriptions of specific locations. The resulting corpus contains almost 7000 individual texts and around 8 million words. By way of comparison, Bieling (2014) built a rich corpus of 42 short stories using more traditional participative methods. To demonstrate transferability we repeated the first two steps of our overall workflow (Fig. 1) for another national park in England – the Broads – since its geographical characteristics deviate strongly from the ones of the Lake District. The Broads is a flat region in the East of England with an exceptionally developed navigable network of rivers and lakes used for sailing. Using entries of the UK national mapping agency gazetteer spatially located within the borders of the Broads from which we removed entries referencing to farms and houses (total of 199 unique entries) and 26 entries from TripAdvisor we extracted 40,402 unique URLs excluding blocked sites (step 2, Fig. 1). Thus, we are confident that our approach is applicable to different landscapes.

Our corpus is well distributed across the whole Lake District and shows a strong relationship to the initial distribution of search terms, suggesting that customising lists would enable us to return more documents. Customisation brings us to a first important implication for landscape research more generally. Methods seeking to classify text remain dependent on lexicons and training on domain specific texts is essential for the development of new methods. As applications of text analysis in landscape research grow, there is a need to develop customised resources and methods for landscape research. We emphasise that our methods have been developed and trained on data in the English language, and we do not expect all results to be culturally invariant (Mark & Turk, 2017). Our texts are produced by a self-selecting group of individuals, who reflect neither all experiences nor opinions about the Lake District. Language itself is biased towards positivity

Table 5
Summary of key questions posed and feedback from expert discussion.

Question	Feedback
Strengths:	
What do you see as particular strengths of this approach and the results presented?	The methods are *repeatable and automatic*, making *comparison*, for example, between each 5-years periods *possible*.
	The *value* of the approach for *Landscape Character Assessment* was *explicitly stated*.
	Possibility to *change search terms* makes *method robust*.
Weaknesses:	Sources may be *biased* to *positive descriptions*.
What weaknesses do you see with respect to the approach and the results presented?	*Missing temporal information* is very important for *monitoring* and estimating *differences due to seasonality*. *Peace and tranquillity* may be dependent on *(unknown) background* of writer. Descriptions are not *stratified* according to either *activity* or *experience* (e.g. looking at versus visiting the summits)
Opportunities:	Monitoring of *landscape quality* and relation to *perception of visitors*. Monitoring of *opinions* towards *access management* actions and other *planning decisions* (e.g., before/after). Identification of *topics and species* (e.g., alpine plants on Scafell), which are *not perceived*, but are *important ecologically*. Such topics could be used in *visitor education*. Incorporating *other sources* (e.g., Twitter, Facebook groups) could show *more negative and instantaneous opinions* (e.g., traffic jams during the Bank Holidays).
How could your organisation use this approach and these results, if at all?	
Threats:	
What dangers do you see in adopting these methods, and in working with these results?	*Potential misrepresentation* of *certain groups* of users (e.g., hill climbers).

(Dodds et al., 2014), and individuals are more likely to report on positive experiences in landscapes (Taylor, Czarnowski, & Flick, 1995). However, by comparing what is reported across a region, we can still make an important contribution to understanding landscape perception. Finally, since we rely on scraping of content, we note the importance of considering ethical issues in so doing (Zimmer, 2018).

Our second contribution builds on our corpus, to analyse and classify different forms of perception (sight, sound and smell) at different scales and using a range of methods. We transferred a random forest classifier, which identified different sound experiences, directly to these texts, demonstrating that our approach is robust. However, we also had to rely on manual annotation to classify rare descriptions (such as those related to smell perception). Despite the emergence of other approaches which classify potential perception based purely on the existence of emitters (e.g., Quercia & Schifanella, 2015), we focussed on experienced (and not potential) perception. Our workflow, extracted and classified 28,445 descriptions referring to sight, 1480 descriptions sound experiences and 78 describing smells. This information is valuable at a range of scales, for example, allowing us to characterise and compare the nature of tranquillity in LCA areas. Our methods rely on linking perception to locations and coordinates, where improvements are still required. For example, we can only separate visited from seen locations in a rudimentary way (c.f. Moncla, Renteria-Agualimpia, Nogueras-Iso, & Gaio, 2014) and we treat all landscape elements as point locations.

The European Landscape Convention, and our expert group, emphasise the importance of landscape as perceived by people, and in turn, we need to know something about who describes landscape. Although we expected our two lists to capture different sorts of users – 'behavioural insiders' and 'empathetic insiders' – this difference turned out to be less clear cut in practice. Furthermore, in traditional monitoring instruments demographic information such as age, gender, and occupation are considered important (Kienast, Degenhardt, Weilenmann, Wäger, & Buchecker, 2012), and for all of these we have no information. We suggest two possible directions here. First, as with the boom in research on social media, there is a need to reflect on ways of modelling who is active in the landscape, who writes about it, and how we can classify characteristic behavioural patterns (c.f. Komossa, van der Zanden, Schulp, & Verburg, 2018). Second, many of our texts contained detailed information about those producing the content. There is no reason why, with appropriate ethical approval and data protection, that such writers cannot be approached and surveyed to reveal more about those producing such data (c.f. boyd, 2007).

Our last question concerned the potential of our approach. First, we hope that the rich examples we have produced exploring the ways in which the Lake District is perceived demonstrate clearly how texts can be extracted, classified and analysed. We see potential of our approach, for example, for 'landscape biography', as here first-person historical narratives are also important (c.f. Kolen & Renes, 2015) and our extraction techniques can contribute to the plurality of collected stories, since the methods could be adapted to historical written accounts. We did not look specifically at the extraction of memories in our work, but covered the aspect of soundscapes important for identification of historical patterns in landscapes (Kolen, Renes, & Bosma, 2018). Second, we evaluated and discussed our approach with an expert group. Although this group was small, they were able to identify and suggest ways such data could be used in practice, showing that our approach has practical utility. Indeed, for monitoring, we see such texts as an interesting way of predicting potential change. For example, one description explicitly comments on the tranquillity of a hill recently resurveyed with a height of more than 2000 feet (and thus reclassified as a mountain!). An online news report comments, "Miller Moss is not the most exciting hill in the world but it should become a little busier now" (Barnard, Jackson, & Bloomer, 2018). Tracking how this landscape element is described in future writing could illustrate if this prediction comes true, but would also require us to more systematically analyse time. To explore temporal change in our corpus, we used the temporal tagger HeidelTime (Strötgen & Gertz, 2010) to extract references to dates. Fig. 9 shows that almost 50% of the texts we analysed are from 2017 and 2018, suggesting that texts describing first-person perception appear to have a rather short period of existence, pointing to the importance of archiving such material if it is to be used for research in the

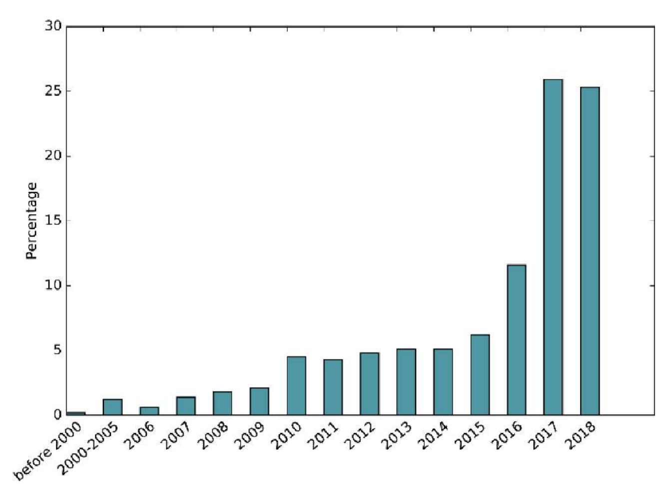


Fig. 9. Temporal distribution of collected texts.

future (Hale, Blank, & Alexander, 2017).

Of course, our approach cannot replace traditional approaches to LCA. Rather we see it as a way of exploring, across large volumes of texts, the diverse ways in which landscape is experienced, and providing impetus to other methods, which seek to better incorporate such perception in LCA and similar approaches.

5. Conclusions and further work

Our starting point was the need to integrate perception, as experienced by those visiting a landscape, in landscape characterisation and monitoring. Inspired by the narrative nature of LCA, and the importance of incorporating different viewpoints and senses, we used the internet as a source to extract and analyse texts capturing first-person perception in the Lake District. Our results demonstrate that, it is possible to build a large, diverse corpus of first-person landscape experiences, and analyse it with respect to multiple senses. More profoundly, they demonstrate that text is a rich, though as yet rarely exploited potential source of landscape information. To utilise such information there is a need to develop landscape specific methods for analysis, which are culturally and linguistically sensitive, and to rethink over-simplistic taxonomies of landscape use. In our texts we find multiple, intertwined experiences, which cannot be meaningfully disentangled into tourist and local perspectives. Text also lays bare the influences of other discourses on the ways in which landscapes are experienced, and reaffirms the importance of considering how guidebooks, modern and historic nature writing, and poetry can influence ways in which landscapes are perceived and remembered (Prior, 2017).

Analysing such information requires that we develop effective approaches to identifying both widely shared views, and also more marginalised opinions, which may capture groups not well represented in the underlying data. Equally, our rich corpora would lend itself to a wide range of other qualitative and quantitative analyses, for example, exploring sentiment with respect to landscape, or perception related to biodiversity indicators.

CRediT authorship contribution statement

Olga Koblet: Conceptualization, Methodology, Software, Validation, Visualization. **Ross S. Purves:** Conceptualization, Methodology.

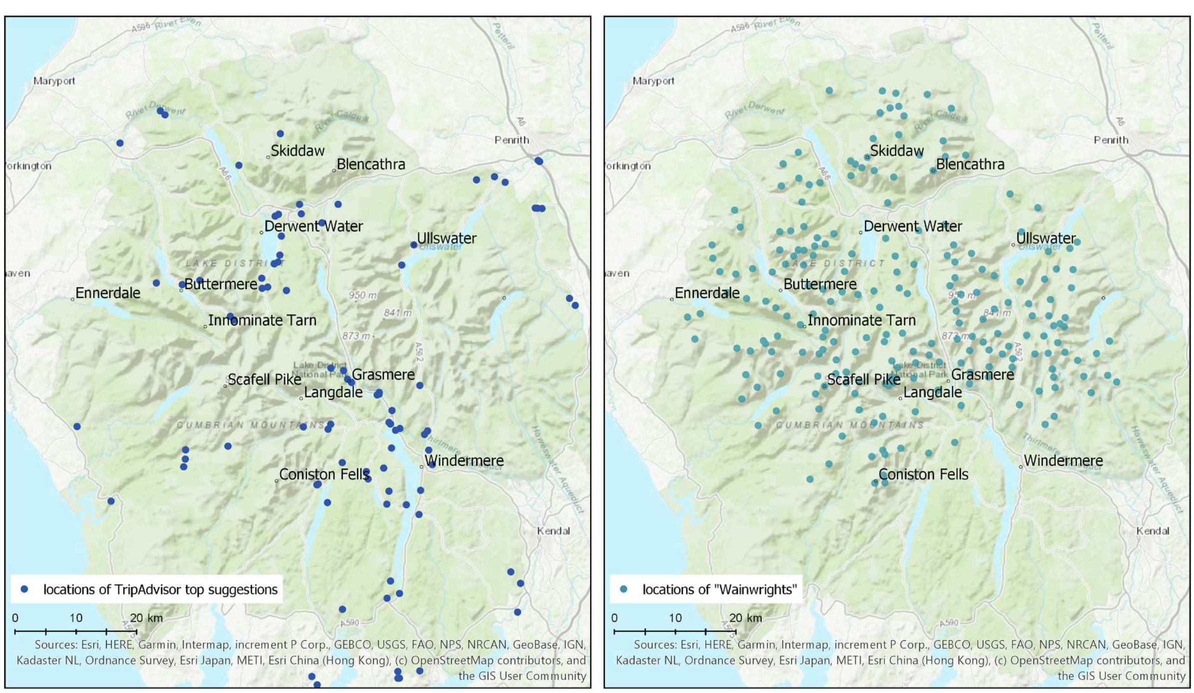
Acknowledgement

We would like to thank all the contributors to ScenicOrNot (Open Database Licence) and Geograph British Isles (Creative Commons Attribution-ShareAlike 2.5 Licence), local experts for their valuable feedback and the reviewers for their insightful comments, which improved the clarity of the article. RSP gratefully acknowledges support from the Swiss National Science Foundation Project EVA (200021E-166788) and the University Research Priority Programme on Language and Space.

Appendix

Terms used for the initial filtering of the returned URLs:

Wikipedia,wikipalapp,wiktionary,facebook,weather,.gov.,panorama,cottage,hotel,for-sale,airbnb,booking,expedia,bustimes,books,pubguide, citypopulation,laterooms,indeed.co.uk, youtube,inn,rentals,yr.no,ordnancesurvey,bedandbreakfast,news,bbc,transportation,streetviewmaps,campsites,nurseries,mypub,forecast,prices,selfcatering,home,countrysideclassroom,geog.port.ac.uk,lakelandcampingbarns,fortune,money,business, observer,shop,cafe,scandal,store,twitter,forbes,linkedin,theguardian,finestproperties,publications,naturalengland,distillery,restaurant,dictionary, highclose,house,availability,onthemarket,windguru,colinday,rooms,aqua3,ferries,property,wikimedia,delivery,ancestry,holidaylettings, accommodation,jobs,online,brew,House,Hotel,tate.org.uk,playdale,finance,solutions,amazon,imdb,thetimes,nytimes,telegraph.co.uk,independent,



Appendix 1. Temporal distribution of collected texts. Left: locations of the search terms from the 'TripAdvisor list'. Right: locations of the search terms from the 'Wainwright list'.

Table Appendix 1
Rules and important distinguishing marks for annotation of first-person perception training data.

First-person landscape descriptions	Not relevant descriptions
• explicit descriptions of perception (Heard Snipe piping for the first time this year; the heather smells lovely; the scent of wet peat and sun-warmed bog myrtle) • events that have already happened as opposed to anticipated ones • descriptions using verbs of motion in combination with personal pronouns 'I' or 'we' (we went to …; I walked 12 miles) • potentially contain references to time (today; Wednesday; this lovely spring morning) • potentially contain descriptions of weather (it was still raining; the sun was shining) • potentially include names of the fellow travelers	• describe anticipated events (next week we go to the magnificent Aira Force waterfall) • present a consistent use of passive voice (it can be done by both car and on foot) • contain imperatives (keep on the road, head north) • contain information to help navigation (you reach; the river on your right) • include lists of walks (five best walks in the Lake District) • are indoors descriptions • weather forecasts • official parish information

Table Appendix 2
Taxonomy of tranquillity for annotation.

Classes of tranquillity	Description
Combination of sight and sound perception	Descriptions, where visual attributes of the scene are as important as sound, and where absence of sounds is implicit (e.g., 'A remembrance service is held here every year and I can't think of a more beautiful and peaceful place to reflect.')[1]
Contrasting sounds	Descriptions reflecting ephemerality of tranquillity by comparing it to other less tranquil locations, different time of day or mentioning sounds which add or detract from overall tranquillity (e.g., 'A moment of peace at Ashness Bridge – rare moments indeed!'[2])
No-movement	Explicit mention of a lack of movement with implied silence and tranquillity (e.g., 'Below to the west Buttermere appeared mirror calm, the blue of the sky reflected deeply in its chill waters.'[3])
Total silence and tranquil sounds	Either descriptions of tranquil sounds without contrast or explicit descriptions of complete silence (e.g., 'Further to the north Blencathra and Skiddaw put in an appearance in the evening sun, and we stopped to listen to the silence – not a sound – very peaceful and relaxing.'[4])

[1] http://juliahedges.blogspot.com/2018/06/a-walk-up-mighty-great-gable.html.
[2] http://www.lakedistrict-walks.co.uk/2016/September/10.09.2016_Bleaberry_Fell.html.
[3] http://www.david-forster.com/section278539_221484.html.
[4] http://www.ramblingpete.walkingplaces.co.uk/day/lakes/martindale.htm.

filmclub,realestate,resort,medical,trains,yellowpages,checkmypostcode,geograph.org.uk,britishplacenames,office,fivestar,research,quora.com, product,clinic,agency,police,science,street,ebay,etsy,paranormal,massage,pinterest,.fr,.au,.jp,.edu,timeanddate,obituaries,lyrics,stackexchange, dailypost,lawyers,washingtonpost,glamour,movies,usatoday,.cnn.,denverpost,metro.co.uk,thesun.co.uk,wiki,scientificamerican,person, startribune,religion,tvtropes,resetera,soundcloud,rottentomatoes,itunes,nypost,britannica,salvationarmy,psychologytoday,cancer,edinburghfestival, vancouversun,spotify,goodreads,foxsports,nbcsports,seattletimes,encyclopedia,dailystar,biography,tvguide,zhidao.baidu,flickr,accident,incident, incidents,cars,gamespot,whitepages,vimeo,startrek,lodging-world,the-saleroom,sun-up.co.uk

Keywords sound:

Babble,bang,beat,beep,blare,blast,boom,bubble,burble,burr,chime,chink,chir,chug,clack,clang,clank,clap,clash,clatter,cling,clink,clomp, clump,clunk,crack,crackle,crash,creak,crepitate,crunch,cry,ding,dong,explode,fizz,fizzle,groan,gurgle,hum,jangle,jingle,knell,lilt,moan,murmur, patter,peal,ping,plink,plonk,plop,plunk,pop,putter,rap,rasp,rattle,ring,roll,rumble,rustle,shriek,shrill,sizzle,splash,splutter,sputter,squelch, strike,swish,swoosh,thrum,thud,thump,thunder,thunk,tick,ting,tinkle,toll,toot,tootle,- trumpet,twang,ululate,vroom,wheeze,whine,whir,whish,whoosh,whump,zing,baa,bark,bay,bellow,blat,bleat,bray,buzz,cackle,call,caw,chatter, cheep,chirp,chirrup,chitter,cluck,coo,croak,crow,cuckoo,drone,gobble,growl,grunt,hee,haw,hiss,honk,hoot,howl,meow,mew,moo,neigh,oink,peep, pipe,purr,quack,roar,scrawk,scream,screech,sing,snap,snarl,snort,snuffle,squawk,squeak,squeal,stridulate,trill,tweet,wail,warble,whimper,whinny, whistle,woof,yap,yell,yelp,yip,yowl,din,echo,resonate,resound,sound,listen,hear,clamor,shout,holler,noise,brawl,discord,grate,gnash, grind,slam,stamp,surd,clonk,blether,blither,ripple,guggle,brattle,chirr,clangor,grumble,whoop,boisterous,shrill,silent,melodic,clamorous, melodious,croaky,muffled,soundless,discordant,squeaky,noiseless,earsplitting,noisy,tacit,thundering,gruff,thunderous,quiet,rasping, tuneful,raspy,raucous,resonant,vociferous,hoarse,rowdy,husky,creaky,loud,screaming,screechy,deafening,hushed,inaudible,stillness,muteness,still, silentness,soundlessness,noiselessness,flick,whisper,susurration,mutter,tranquillity,tranquility,peace,restfulness,quietness,calm,calmness,quietude, serenity,peacefulness,reposefulness,shush,whisht,whiffle,mute,tacitly,quietlike,mouselike,tongueless,dumb,mousy,whisperless,voiceless,halcyon, peaceful,peaceable,restful,tranquil,undisturbed

Keywords smells:

Reek,smell,stink,inhale,aroma,odour,scent,malodor,malodour,stench,fetor,mephitis,acridity,aroma,fragrance,acrid,antiseptic,foetid,fetid,fragrant, musky,musty,noxious,whiffy,odorous,pungent,putrid,rancid,scentless

Scenic pairs:

('middle','distance'),('far','distance'),('steep','slopes'),('highest','point'),('lower','slopes'),('low','tide'),('long','ridge'),('small','loch'),('fine','view'), ('northern','slopes'),('distant','view'),('small','lochan'),('high','point'),('south','ridge'),('small','cairn'),('western','slopes'),('west','ridge'),('small','island'), ('southern','slopes'),('prominent','hill'),('south','shore'),('rough','grazing'),('east','ridge'),('far','side'),('small','islands'),('distant','hill'),('steep','slope'), ('high','tide'),('natural','arch'),('surrounding','hills'),('southern','side'),('south','coast'),('eastern','slopes'),('rocky','outcrops'),('eastern','side'),('small','hill'), ('prominent','peak'),('small','crags'),('freshwater','loch'),('northern','shore'),('steep','ridge'),('eastern','top'),('fine','views'),('superb','views'),('highest','- hill'),('large','boulder'),('great','views'),('good','views'),('good','view'),('gentle','slopes'),('sharp','peak'),('moderate','slopes'),('rough','grass'),('remote','- hill'),('coastal','scenery'),('southern','ridge'),('conical','hill'),('narrow','ridge'),('deep','gorge'),('broad','ridge'),('near','distance'),('west','coast'), ('rough','moorland'),('unnamed','lochan'),('lower','part'),('magnificent','views'),('good','path'),('cliff','top'),('south','end'),('upper','part'),('great','view'),('flat','summit'),('distant','views'),('similar','view'),('steep','valley'),('rocky','spur'),('rough','track'),('true','summit'),('highest','peak'),('rocky','promontory'),('lower','top'),('old','pier'),('fresh','snow'),('rocky','hill'),('lewisian','gneiss'),('huge','boulder'),('main','summit'),('exposed','rock'),('west','shore'),('rocky','summit'),('beautiful','beach'),('northwest','ridge'),('big','hill'),('northern','side'),('southwest','ridge'),('fine','viewpoint'),('cliff','edge'),('rough','slopes'),('northern','ridge'),('small','lochans'),('little','hill'),('low','point'),('south','corner'),('western','side'),('brown','trout'),('northern','tip'),('far','shore'),('high','ground'),('shallow','loch'),('eastern','shore'),('southern','tip'),('tussocky','grass'),('small','bay'),('coastal','path'),('small','waterfall'),('southern','shore'),('steep','descent'),('sandy','beach'),('few','places'),('stepping','stones'),('many','summits'),('more','rocks'),('rocky','pavement'),('rocky','bit'),('shapely','peak'),('cambrian','quartzite'),('sharp','ridge'),('topped','mountain'),('small','trout'),('horizontal','strata'),('early','snow'),('low','hill'),('north','slopes'),('dissected','bog'),('long','loch'),('largest','loch'),('highest','mountain'),('beautiful','scenery'),('superb','view'),('top','end'),('rocky','slopes'),('tiny','lochan'),('made','path'),('eastern','part'),('deep','pool'),('good','viewpoint'),('rocky','coastline'),('unnamed','top'),('tidal','island'),('little','snow'),('rough','hill'),('rocky','coast'),('steep','ascent'),('wonderful','view'),('wonderful','views'),('west','face'),('small','hills'),('lowest','point'),('upper','valley'),('covered','slopes'),('southeast','side'),('spectacular','view'),('north','end'),('clear','view'),('south','top'),('small','outcrop'),('steep','drop'),('stunning','views'),('unnamed','hill'),('rocky','beach'),('rocky','ridge'),('deep','snow'),('right','side'),('shaped','valley'),('right','skyline'),('heavy','snow'),('left','skyline'),('boggy','ground'),('glen','floor'),('gentle','ridge'),('flat','floor'),('facing','slopes'),('faint','path'),('far','horizon')

Unattractive pairs:

('industrial','estate'),('dual','carriageway'),('new','housing'),('new','development'),('small','estate'),('residential','area'),('new','estate'),('industrial','units'),('terraced','houses'),('slip','road'),('former','airfield'),('new','building'),('large','estate'),('cooling','towers'),('many','buildings'),('detached','houses'),('modern','estate'),('busy','junction'),('modern','housing'),('new','centre'),('retail','park'),('new','developments'),('modern','building'),('small','development'),('new','station'),('former','factory'),('recent','development'),('main','runway'),('industrial','area'),('northbound','carriageway'),('new','buildings'),('large','complex'),('former','village'),('suburban','road'),('high','voltage'),('multi','storey'),('new','park'),('major','road'),('terraced','housing'),('major','junction'),('central','reservation'),('wartime','airfield'),('local','shops'),('residential','road'),('detached','housing'),('former','garage'),('filling','station'),('large','park'),('retail','outlet'),('large','village'),('large','roundabout'),('industrial','unit'),('new','units'),('large','development'),('large','centre'),('typical','housing'),('closed','pub'),('new','part'),('small','businesses'),('local','centre'),('perimeter','fence'),('new','stadium'),('new','estates'),('single','carriageway'),('adjacent','site'),('carriageway','road'),('new','homes'),('new','block'),('new','turbines'),('new','roundabout'),('current','building'),('old','a1'),('overflow','park'),('main','office'),('back','street'),('social','club'),('chinese','takeaway'),('flat','roofs'),('industrial','premises'),('main','station'),('new','apartments'),('more','housing'),('old','machinery'),('named','road'),('residential','street'),('solar','panels'),('private','houses'),('local','road'),('special','train'),('general','store'),('typical','houses'),('public','houses'),('near','junction'),('local','office'),('old','works'),('staggered','junction'),('new','lease'),('old','base'),('heavy','industry'),('modern','style'),('industrial','use'),('heavy','plant'),('new','store'),('slip','roads'),('new','blocks'),('main','stand'),('new','hall'),('main','hall'),('typical','development'),('several','streets'),('busy','roundabout'),('many','businesses'),('high','tension'),('suburban','street'),('new','premises'),('main','carriageway'),('large','works'),('agricultural','produce'),('crossing','gates'),('organic','farm'),('parked','cars'),('industrial','site'),('much','traffic'),('underground','reservoir'),('overhead','lines'),('new','construction'),('old','factory'),('multiple','unit'),('rubbish','tip'),('terraced','cottages'),('adjacent','park'),('residential','areas'),('static','caravans'),('old','hospital'),('overhead','cables'),('new','hospital'),('main','centre'),('new','extension'),('staggered','crossroads'),('major','development'),('unusual','design'),('next','station'),('large','sheds'),('large','hospital'),('busy','intersection'),('free','house'),('many','estates'),('residential','estate'),('new','facilities'),('wartime','factory'),('large','tower'),('underground','workings'),('former','depot'),('large','plant'),('new','flats'),('concrete','building'),('motorway','junction'),('current','station'),('large','barns'),('industrial','complex'),('striking','building'),('main','industry'),('modern','centre'),('local','service'),('guided','busway'),('multi','purpose'),('tall','buildings'),('double','glazing'),('suburban','housing'),('light','units'),('high','density'),('many','stations'),('retail','centre'),('nearby','line'),('largest','centre'),('agricultural','equipment'),('executive','houses'),('main','hospital'),('electric','pumps'),('mini','roundabout'),('closed','station'),('bound','carriageway'),('new','build'),('other','facilities'),('large','quarry'),('substantial','buildings'),('industrial','estates'),('built','houses'),('busy','carriageway'),('rolling','stock'),('large','blocks'),('former','estate'),('small','site'),('last','building'),('front','wall'),('agricultural','machinery'),('big','houses'),('large','buildings'),('suburban','development'),('main','gates'),('former','works'),('old','runways'),('built','building'),('large','range'),('old','runway'),('more','buildings'),('rural','lane'),('integral','part'),('third','rail'),('second','bridge'),('tidy','farm'),('arterial','road'),('main','lines'),('mobile','homes')

List of British mammals:

Beaver,vole,mouse,rat,dormouse,squirrel,porcupine,hare,rabbit,mole,shrew,hedgehog,bat,pipistrelle,dog,fox,seal,walrus,marten,weasel,polecat,otter,badger,wildcat,cat,mink,coati,boar,goat,sheep,cattle,deer,reindeer,moose,muntjac,buffalo,whale,dolphin,beluga,porpoise,orca,cow,stag,cattle,lamb

List of natural elements:

Water,river,tree,beach,sea,snow,coast,stone,rain,grass,harbour,seaside,leaves,lake,wood,plant,sand,pond,mist,fog,ice,rock,forest,hill,island,leaf,mountain,bay,waterfall,loch,wave,seafront,mud,landscape,summit,valley

Annotated data of Geograph: https://github.com/olgaches/Geograph_sound_descriptions.

References

Arias, F. J. C. (2019). Fuzzy String Matching in Python [WWW Document]. URL: https://www.datacamp.com/community/tutorials/fuzzy-string-python.

Barnard, J., Jackson, G., & Bloomer, J. (n.d.) England has a new mountain: Miller Moss. Now go find it [WWW Document]. 2018. URL: https://www.grough.co.uk/magazine/2018/08/09/england-has-a-new-mountain-miller-moss-now-go-find-it (accessed 06.10.19).

Bieling, C. (2014). Cultural ecosystem services as revealed through short stories from residents of the Swabian Alb (Germany). *Ecosystem Services, 8*, 207–215.

Bieling, C., Plieninger, T., Pirker, H., & Vogl, C. R. (2014). Linkages between landscapes and human well-being: An empirical exploration with short interviews. *Ecological Economics, 105*, 19–30.

Bing Web Search [WWW Document]. (2019). URL: https://azure.microsoft.com/en-us/services/cognitive-services/bing-web-search-api/ (accessed 10.02.19).

boyd, d. (2007). Why youth (heart) Social network sites: The role of networked publics in teenage social life. In D. Buckingham (Ed.). *MacArthur foundation series on digital learning – Youth, identity, and digital media volume*. Cambridge, MA: MIT Press.

Brown, G., & Reed, P. (2009). Public participation GIS: A new method for use in national forest planning. *Forest Science, 55*, 166–182.

Bruns, D., & Stemmer, B. (2018). Landscape assessment in Germany. *Routledge handbook of landscape character assessment* (pp. 154–167). .

Butler, A. (2016). Dynamics of integrating landscape values in landscape character assessment: The hidden dominance of the objective outsider. *Landscape Research, 41*, 239–252.

Carles, J. L., Barrio, I. L., & De Lucio, J. V. (1999). Sound influence on landscape values. *Landscape and Urban Planning, 43*, 191–200.

Caspersen, O. H. (2009). Public participation in strengthening cultural heritage: The role of landscape character assessment in Denmark. Geogr. Tidsskr. – Danish. *Journal of Geography, 109*, 33–45.

Chesnokova, O., & Purves, R. S. (2018a). *Automatically creating a spatially referenced corpus*

of landscape perception. 12th ACM SIGSPATIAL workshop on geographic information retrieval. Seattle, WA, USA: ACM.
Chesnokova, O., & Purves, R. S. (2018b). From image descriptions to perceived sounds and sources in landscape: Analyzing aural experience through text. *Applied Geography, 93*, 103–111.
Chesnokova, O., Taylor, J. E., Gregory, I. N., & Purves, R. S. (2019). Hearing the silence: finding the middle ground in the spatial humanities? Extracting and comparing perceived silence and tranquillity in the English Lake District. *International Journal of Geographical Information Science, 33*, 2430–2454.
Clemetsen, M., Krogh, E., & Thorén, K. H. (2011). Landscape perception through participation: Developing new tools for landscape analysis in local planning processes in Norway. In M. Jones, & M. Stenseke (Eds.). *The European landscape convention. Challenges of participation* (pp. 219–237). Springer.
Coates, P. A. (2005). The strange stillness of the past: Toward an environmental history of sound and noise. *Environmental History Durh. N. C. 10*, 636–665.
Council of Europe. (2000). European landscape convention. Rep. Conv. Florence ETS No. 17, 8.
Criminisi, A., Shotton, J., & Konukoglu, E. (2011). Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Foundations and Trends in Computer Graphics and Vision, 7*, 81–227.
Cronon, W. (1992). A place for stories: Nature, history, and narrative. *Journal of American History, 78*, 1347–1376.
Dann, G. M. S., & Jacobsen, J. K. S. (2003). Tourism smellscapes. *Tourism Geography, 5*, 3–25.
Dara-Abrams, D. (2011). Jenks natural breaks [WWW Document]. URL: https://gist.github.com/drewda/1299198.
Daume, S., Albert, M., & von Gadow, K. (2014). Forest monitoring and social media – Complementary data sources for ecosystem surveillance? *Forest Ecology and Management, 316*, 9–20.
Davies, C. (2013). Reading geography between the lines: Extracting local place knowledge from text. *Lecture Notes in Computer Science (including its subseries Lecture Notes in Artificial Intelligence (LNAI) and Lecture Notes in Bioinformatics), 8116 LNCS*, 320–337.
de Kunder, M. (2019). The size of the World Wide Web (The Internet) [WWW Document]. URL: https://www.worldwidewebsize.com/ (accessed 10.01.19).
Dodds, P. S., Clark, E. M., Desu, S., Frank, M. R., Reagan, A. J., Williams, J. R., ... Danforth, C. M. (2014). Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences, 112*, 2389–2394.
Donaldson, C., Gregory, I. N., & Taylor, J. E. (2017). Locating the beautiful, picturesque, sublime and majestic: Spatially analysing the application of aesthetic terminology in descriptions of the English Lake District. *Journal of Historical Geography, 56*, 43–60.
Edwards, J. (2018). Literature and sense of place in UK landscape strategy. *Landscape Research, 44*, 659–670.
Gonzalez, J., Rodrigues, P., & Cohen, A. (2017). Fuzzywuzzy: Fuzzy string matching in python [WWW Document]. URL: https://github.com/seatgeek/fuzzywuzzy.
Esri. (2019). ArcGIS API for Python [WWW Document]. URL: https://pro.arcgis.com/en/pro-app/arcpy/get-started/arcgis-api-for-python.htm.
Fairclough, G., Sarlöv Herlin, I., & Swanwick, C. (Eds.). (2018). *Routledge handbook of landscape character assessment*. Routledge.
Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database.* MA: MIT Press.
Fisher, J. A. (1999). The value of natural sounds. *Journal of Aesthetic Education, 33*, 26–42.
Galaz, V., Crona, B., Daw, T., Bodin, Ö., Nyström, M., & Olsson, P. (2010). Can web crawlers revolutionize ecological monitoring? *Frontiers in Ecology and the Environment, 8*, 99–104.
Granö, J. G. (1997). *Pure geography.* John Hopkins University Press.
Greenaway, M. (2017). Web-scraping policy [WWW Document]. URL: https://www.ons.gov.uk/aboutus/transparencyandgovernance/lookingafterandusingdataforpublicbenefit/policies/policieswebscrapingpolicy (accessed 07.01.19).
Hale, S. A., Blank, G., & Alexander, V. D. (2017). Live versus archive: Comparing a web archive and to a population of webpages. In N. Brügger, & R. Schroeder (Eds.). *The web as history* (pp. 45–61). London: UCL Press.
Herlin, I. S. (2016). Exploring the national contexts and cultural ideas that preceded the Landscape Character Assessment method in England. *Landsc. Res. 41*, 175–185.
Hewlett, D., Harding, L., Munro, T., Terradillos, A., & Wilkinson, K. (2017). Broadly engaging with tranquillity in protected landscapes: A matter of perspective identified in GIS. *Landsc. Urban Plann. 158*, 185–201.
Honnibal, M., & Johnson, M. (2015). An improved non-monotonic transition system for dependency parsing 1373–1378.
Jockers, M. (2013). *Macroanalysis: Digital methods and literary history.* University of Illinois Press.
Joho, H., & Sanderson, M. (2000). Retrieving descriptive phrases from large amounts of free text. *Proceedings of the ninth international conference on information and knowledge management (CIKM)* (pp. 180–186). USA: Mclean.
Jones, C. B., Purves, R. S., Clough, P. D., & Joho, H. (2008). Modelling vague places with knowledge from the Web. *International Journal of Geographical Information Science, 22*, 1045–1065.
Jones, M., & Stenseke, M. (Eds.). (2011). *The European landscape convention. Challenges of participation.* Springer.
Kaji, N., & Kitsuregawa, M. (2007). Building lexicon for sentiment analysis from massive collection of HTML documents. *EMNLP-CoNLL. Prague* (pp. 1075–1083). .
Kienast, F., Degenhardt, B., Weilenmann, B., Wäger, Y., & Buchecker, M. (2012). GIS-assisted mapping of landscape suitability for nearby recreation. *Landscape and Urban Planning, 105*, 385–399.
Kienast, F., Frick, J., van Strien, M. J., & Hunziker, M. (2015). The Swiss landscape monitoring program – A comprehensive indicator set to measure landscape change. *Ecological Modelling, 295*, 136–150.
Kienast, F., Wartmann, F., Zaugg, A., & Hunziker, M. (2019). *A Review of Integrated Approaches for Landscape Monitoring*.
Kolen, J., & Renes, J. (2015). Landscape biographies: Key issues. In J. Kolen, H. Renes, & R. Hermans (Eds.). *Landscape biographies* (pp. 21–48). Amsterdam University Press.
Kolen, J., Renes, H., & Bosma, K. (2018). The landscape biography approach to landscape characterisation. *Routledge handbook of landscape character assessment* (pp. 168–184). .
Komossa, F., van der Zanden, E. H., Schulp, C. J. E., & Verburg, P. H. (2018). Mapping landscape potential for outdoor recreation using different archetypical recreation user groups in the European Union. *Ecological Indicators, 85*, 105–116.
Krause, B. (2008). Anatomy of the soundscape. *Journal of the Audio Engineering Society,* 56.
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174.
Lawson, R. (2015). *Web scraping with Python.* Packt Publishing Ltd.
Leidner, J. L., Sinclair, G., & Webber, B. (2003). Grounding spatial named entities for information extraction and question answering. *Proceedings of the HLT-NAACL 2003 workshop on analysis of geographic references. Stroudsburg, PA, USA* (pp. 31–38). .
Levin, B. (1993). *English verb classes and alternations.* University of Chicago Press.
Liu, B. (2012). *Sentiment analysis and opinion mining.* Toronto: Morgan & Claypool Publishers.
Lu, Y., Castellanos, M., Dayal, U., & Zhai, C. (2011). Automatic construction of a context-aware sentiment lexicon: An optimization approach. *WWW 2011 – Session: semantic analysis. Hyderabad, India* (pp. 347–356). .
Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods, 41*, 558–564.
MacFarlane, R., Haggett, C., Fuller, D., Dunsford, H., & Carlisle, B. (2004). Tranquillity mapping: Developing a robust methodology for planning support.
Majid, A., & Burenhult, N. (2014). Odors are expressible in language, as long as you speak the right language. *Cognition, 130*, 266–270.
Manning, C. D., & Schutze, H. (1999). *Foundations of statistical natural language processing.* The MIT Press.
Mark, D. M., & Turk, A. G. (2017). Ethnophysiography. *International Encyclopedia of Geography: People, the Earth, Environment and Technology,* 1–11.
Moncla, L., Gaio, M., & Mustière, S. (2014). Automatic itinerary reconstruction from texts. *Eighth international conference on geographic information science (GIScience 2014)* (pp. 253–267). .
Moncla, L., Renteria-Agualimpia, W., Nogueras-Iso, J., & Gaio, M. (2014). *Geocoding for texts with fine-grain toponyms: An experiment on a geoparsed hiking descriptions corpus. Proceedings of the 22nd ACM SIGSPATIAL international conference on advances in geographic information systems. Dallas/Fort Worth, TX, USA*.
Nielsen, J. (2006). The 90-9-1 rule for participation inequality in social media and online communities [WWW Document]. URL: http://www.useit.com/alertbox/participation_inequality.html.
Nomination. (n.d.) Nomination of the English Lake District for inscription on the world heritage list. [WWW Document]. 2017. URL: https://whc.unesco.org/en/list/422 (accessed 07.09.19).
Overell, S., & Rüger, S. (2008). Using co-occurrence models for placename disambiguation. *International Journal of Geographical Information Science, 22*, 265–287.
Palmer, C., & Brady, E. (2007). Landscape and value in the work of Alfred Wainwright (1907–1991). *Landscape Research, 32*, 397–421.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Blondel, M. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research,* 2825–2830.
Pheasant, R., Horoshenkov, K., Watts, G., & Barrett, B. (2008). The acoustic and visual factors influencing the construction of tranquil space in urban and rural environments tranquil spaces-quiet places? *Journal of the Acoustical Society of America, 123*, 1446–1457.
Pheasant, R. J., & Watts, G. R. (2015). Towards predicting wildness in the United Kingdom. *Landscape and Urban Planning, 133*, 87–97.
Prior, J. (2017). Sonic environmental aesthetics and landscape research. *Landscape Research, 42*, 6–17.
PostGIS. (2019). DBScan Clustering [WWW Document]. URL: https://postgis.net/docs/ST_ClusterDBSCAN.html.
Quercia, D., & Schifanella, R. (2015). *Smelly maps: The digital life of urban smellscapes. 9th international AAAI conference on web and social media. Oxford, UK*.
Rattenbury, T., & Naaman, M. (2009). Methods for extracting place semantics from Flickr tags. *ACM Transactions on the Web, 3*, 1–30.
Relph, E. (1976). *Place and placelessness.* London: Pion Press.
Richards, D. R., & Tunçer, B. (2018). Using image recognition to automate assessment of cultural ecosystem services from social media photographs. *Ecosystem Services, 31*, 318–325.
San Roque, L., Kendrick, K. H., Norcliffe, E., Brown, P., Defina, R., Dingemanse, M., ... Majid, A. (2015). Vision verbs dominate in conversation across cultures, but the ranking of non-visual verbs varies. *Cognitive Linguistics, 26*, 31–60.
Silberschatz, A., & Tuzhilin, A. (1996). What makes patterns interesting in knowledge discovery systems. *IEEE Transactions on Knowledge and Data Engineering, 8*, 970–974.
Srinivasa-Desikan, B. (2018). *Natural language processing and computational linguistics: A practical guide to text analysis with Python, Gensim, spaCy, and Keras.* Packt Publishing Ltd.
Strötgen, J., & Gertz, M. (2010). HeidelTime: High quality rule-based extraction and normalization of temporal expressions. *5th international workshop on semantic evaluation* (pp. 321–324). Uppsala, Sweden: ACL. Association for Computational Linguistics.
Swanwick, C., & Fairclough, G. (2018). Landscape character: Experience from Britain. In G. Fairclough, I. Sarlöv Herlin, & C. Swanwick (Eds.). *Routledge handbook of landscape character assessment* (pp. 21–36). Routledge.

Taylor, J. E. (2018). Echoes in the mountains: The romantic lake district's soundscape. *Studies in Romanticism, 57*, 383–406.
Taylor, J. G., Czarnowski, K. J., & Flick, S. (1995). The importance of water to Rocky Mountain National Park visitors: An adaptation of visitor-employed photography to natural resources management. *Journal of Applied Recreation Research, 20*, 61–85.
Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järv, O., ... Di Minin, E. (2019). Social media data for conservation science: A methodological overview. *Biological Conservation, 233*, 298–315.
Tudor, C. (2014). An approach to landscape character assessment.
van den Bosch, A., Bogers, T., & de Kunder, M. (2016). Estimating search engine index size variability: A 9-year longitudinal study. *Scientometrics, 107*, 839–856.
Wartmann, F. M., Acheson, E., & Purves, R. S. (2018). Describing and comparing landscapes using tags, texts, and free lists: An interdisciplinary approach. *International Journal of Geographical Information Science, 32*, 1–21.
Watkins, D. (2008). Landscape character assessment and guidelines.
Winter, B., Perlman, M., & Majid, A. (2018). Vision dominates in perceptual language: English sensory vocabulary is optimized for usage. *Cognition, 179*, 213–220.
Zachara, M., & Palka, D. (2016). Comparison of text-similarity metrics for the purpose of identifying identical web pages during automated web application testing. *Information systems architecture and technology: Proceedings of 36th international conference on information systems architecture and technology – ISAT 2015 – Part II* (pp. 25–35). .
Zimmer, M. (2018). Addressing conceptual gaps in big data research ethics: An application of contextual integrity. *Social Media Society, 4*.